IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received June 1, 2018, accepted July 9, 2018, date of publication July 23, 2018, date of current version August 20, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2858260

VISOS: A Visual Interactive System for Spatial-Temporal Exploring Station Importance Based on Subway Data

TAO TANG[1], XIANGJIE KONG[2], (Senior Member, IEEE), MENGLIN LI[2], JINZHONG WANG[2,3], GUOJIANG SHEN[4], AND XINSHUANG WANG[2]

[1]Chengdu College, University of Electronic Science and Technology of China, Chengdu 611731, China
[2]Key Laboratory for Ubiquitous Network and Service Software of Liaoning Province, School of Software, Dalian University of Technology, Dalian 116620, China
[3]School of Management and Journalism, Shenyang Sport University, Shenyang 110102, China
[4]College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China

Corresponding authors : Xiangjie Kong (xjkong@ieee.org) and Guojiang Shen (gjshen1975@zjut.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61572106, in part by the Natural Science Foundation of Liaoning Province, China, under Grant 201602154, and in part by the Fundamental Research Funds for the Central Universities under Grant DUT18JC09.

ABSTRACT In urban cities, multiple intelligent transportation systems generate a large amount of traffic data. Researchers can make well use of these data to provide solutions for solving numerous existing traffic problems, such as traffic congestions and urban transportation resource allocating. Thus, it is meaningful and feasible for traffic researchers to collect these data and analyze the concealed human mobility based on them. In this paper, we propose a visual interactive subway system (VISOS). The system incorporates subway data visualization module, spatial–temporal exploration module, and station clustering module. VISOS utilizes k-means clustering algorithm to explore the subway data interactively, analyze human mobility pattern responsively, and identify functional characteristics of subway stations precisely. In addition, in this paper, we provide a comprehensive spatial–temporal exploration based on the real Shanghai subway data to analyze the importance level of subway stations.

INDEX TERMS Spatial-temporal exploration, station clustering, visual exploration, subway visualization interactive system.

I. INTRODUCTION

In our daily life, urban intelligent transportation systems are embedded into multiple public transport systems such as bus system and rail transit systems. Most of these systems collect diverse large-scale time series data. These data reflect traffic conditions vividly so that they can be used to improve traffic conditions and provide reliable services with scientific analysis [1]. The subway system is a typical urban rail transit system. It provides urban residents a great convenience for daily travel. Thus, the recorded traffic data of subway stations contain plenty of traffic information. For instance, subway smart card transaction data not only indicates passengers' locations but also suggests passengers' originating and destination stations of each passenger. Collecting all these data can reveal the pattern of human mobility. Mining these subway data deeply, we can differentiate the functions of different subway stations [2] and detect urban traffic congestion and anomalies [3]. Regional functions partition and anomalies detection are significant to understand the transportation infrastructure of cities and help to solve the problems in urban transportation systems [4].

However, the raw subway data is not easy for us to discover the useful information directly due to its abstract attributes. Fortunately, with the development of visualization technologies, it is easier for us to analyze these subway data utilizing visualization technology [5]. Meanwhile, Chen *et al.* [6], Gonçalves *et al.* [7], and Zheng *et al.* [8] have described the data visualization techniques in detail, which provides a solid foundation for this paper. Combining visual interface with human computer interaction, visualization technology enables the users collect and analyze data intuitively. Thus, users can modify the model parameters through interactive operations to display corresponding visualization results automatically and directly. After taking

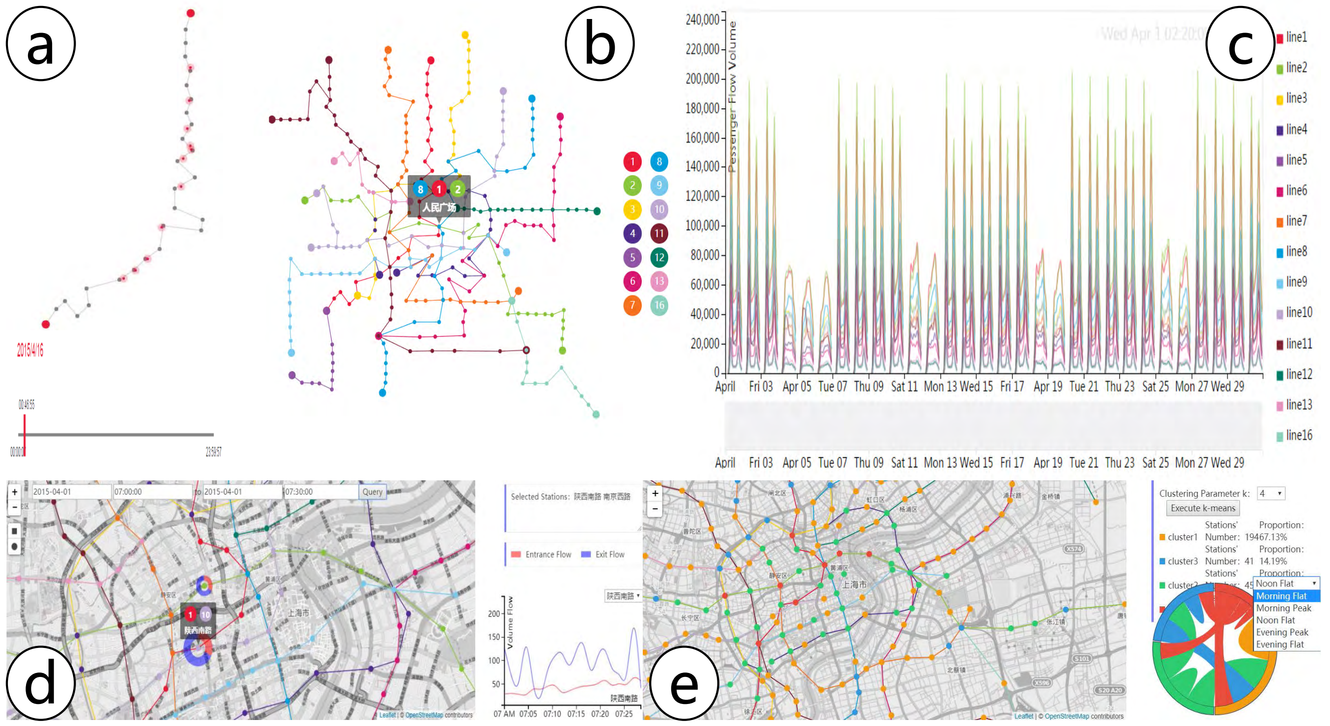


FIGURE 1. **A visual interactive system of subway for the spatial-temporal exploration in human mobility of Shanghai based on subway data. Subway running view (a) provides a dynamic chart for displaying the subway running conditions in subway Line 1. Line and station view (b) supports users interact with the line and station information. Passenger flow view (c) enables users to interact with the time axis for obtaining the specific passenger flows over a month. The spatial-temporal exploration module (d) provides users an efficient interface to query the specific stations' passenger flows by selecting time and space range. The station clustering module (e) enables users to set the parameter of clustering algorithm and implement station clustering.**

visual analysis, users can obtain more valuable information hidden in raw subway data [9].

In this paper, we take an in-depth exploration and mining on Shanghai subway data and the smart card transaction data of Shanghai subway. After that, we propose an interactive system based on Shanghai subway data named VISOS, which combines the clustering algorithm in data mining, aiming to analyze the laws of human mobility. As shown in Fig. 1, VISOS contains three visual views and two functional interface for spatial-temporal exploration and station clustering. In this system, we provide visualization views for visualizing Shanghai subway data more suitably. VISOS enables users to set $k - means$ parameter for station clustering so that they can find the stations which have the similar passenger flow, to analyze the importance of stations and the pattern of human mobility. Users can also query the specific stations' passenger flow by limiting time and space range. Based on the proposed visualization views and the results of station clustering in the system, we verify the effectiveness and practicality of this visualization system by analyzing the real-world subway data.

The main contributions of this paper include:

1) We propose three visualization views based on Shanghai subway datasets: station line view, train running view, and passenger flow view. Our goal is exploring and analyzing Shanghai subway data through these views in an interactive method. Furthermore, we integrate these views into the subway data visualization module of the system.

2) We integrate the spatial-temporal query function into the spatial-temporal exploration module for taking spatial-temporal exploration based on the smart card transaction data of Shanghai subway. The smart card transaction data has spatial and temporal attributes. Therefore, it is necessary to conduct spatial-temporal exploration. Then, combining with spatial-temporal exploration module, we propose station clustering module for clustering the subway stations by implementing K-means algorithm.

3) This paper provides a complete subway visualization system with interactive views such that users can set time and spatial attributes to explore spatial-temporal information in the Shanghai subway data. This system also enables users to explore the regularity of the human mobility in Shanghai subway system.

The remaining part of the paper proceeds as follows: the related work of trajectories research and traffic data visualization is displayed in Section II. Section III lays the design requirements, overall design of our subway station visualization system, and the detailed design of this system. Section IV presents the visualization analysis of human mobility, spatial-temporal exploration, and station clustering results. Finally, this paper concludes in Section V.

II. RELATED WORK

We introduce the relevant work about human mobility exploration and spatial-temporal analysis based on traffic-related data in this section.

A. HUMAN MOBILITY EXPLORATION

With the development of sensor technologies and the spread of multiple mobile devices, a large amount of spatial trajectory data can be generated, these data often contain the records of multitudinous moving objects, such as people, animals, and vehicles. Therefore, it is meaningful for researchers to find the mobility patterns of these moving objects [10]. Rahim *et al.* [11] investigate the future implementation of Vehicular Social Networks and presents a literature review on socially-aware applications of Vehicular Social Networks and mobility modeling. They give an overview of the recommendation systems and route planning protocols. Li *et al.* [12] propose an augmented MapReduce framework on visualization of traffic data, which improves the accuracy within different time scale. Chen *et al.* [13] propose an interactive visual analytics system for exploring the movement patterns based on the sparsely sampled social media data with Geotagging information. This system enables users to use uncertainty model to filter and select reliable data and take spatial-temporal analysis based on these preprocessed data. With the help of this system, users can explore the movement patterns from social media data with Geotagging information. Kong *et al.* [14] propose a novel approach for chronically detecting traffic anomalies based on crowdsourced bus trajectory data. First, they model the traffic conditions by partitioning the raw data into two segments with spatial and temporal attributes. Then they use the anomalous segments to detect the anomalous regions, to further make suggestions for the traffic planning. Wu *et al.* [15] propose an interactive visual analytics system named TelCoVis to help users to take co-occurrence exploration in urban human mobility based telco data by integrating several biclustering techniques. Xia *et al.* [16] extract essential features of human mobility and model the large-scale green urban mobility by mining the big urban traffic data, aiming at evaluating the perception applications for green transportation systems. Besides, they perform simulations for finding the factors which can influence the vehicular mobility. Ning *et al.* [17] propose an efficient and novel heuristic scheme for controling the power of User Equipments and selecting relay, aiming at allocating Resource Blocks following an in-band NB-IoT solution. Al-Dohuki *et al.* [18] propose a semantic-based analysis scheme named SemanticTraj to process the massive taxi trajectory data and integrate the processed results with a set of visualization techniques. This approach not only enables users to query the keywords based on the terms, but also let the users explore the trajectory data interactively through the visualization interface of SemanticTraj. Kong *et al.* [19] present a detailed approach for generating social vehicular mobility dataset from the raw taxi data. After modeling and analyzing the taxi dataset, they further combine official traffic data to describe the road network conditions. Finally, they conduct simulation based on the raw taxi data and the urban functional regions for reproducing the scene by producing the mobility dataset.

B. SPATIAL-TEMPORAL ANALYSIS

Human mobility analysis based on traffic data is a popular topic [20]. Cui *et al.* [21] design a geometry-based edge-clustering framework which is able to cluster edges into bundles and decrease the overall edge crossings. Tominski *et al.* [22] present a novel spatiotemporal approach to visualizing trajectory attribute data and achieve good performance. Ferreira *et al.* [23] propose a visual model with origin-destination query functions based the New York City taxi data that allow users to conduct taxi trips visual query using spatial and temporal attributes. Based on this model, they also integrate a visual system for rendering the query results more efficiently such that users can obtain the hidden details from the raw traffic data. Zeng *et al.* [9] propose two visualization modules named isotime flow map view and the OD-pair journey view for displaying the smart card transaction data and the basic data of Singapore subway and the smart card transaction data of Singapore bus lines, assisting in the analysis of passenger travel patterns. Huang *et al.* [24] propose a visual analytics method named TrahGraph, which uses the road network structure based on Shenzhen taxi trajectory data, aiming at revealing the importance of city streets. TrajGraph simplifies the map model at the regional level by utilizing the graph segmentation algorithm for the road networks. After graph centralization process, users can obtain city traffic patterns by examining the importance of streets in an interactive way. Kristian Kloeckl *et al.*[1] visualize French high-speed rail data using the isochronal charts of the high-speed rail network and high-speed real-time operational maps. This visualization scheme provides a new perspective for displaying the high-speed rail network operation. Yang *et al.* [25] propose a hybrid visualization method named MapTrix, combining the OD matrix and flow map presentations, aiming at displaying regional many-to-many flows. Miranda *et al.* [26] define the concept of the dynamic spatial-temporal activities in a city across multiple temporal resolutions as the ‘‘urban pulse’’. They also propose an visual exploration framework so that users can explore the pulses within and across plenty cities under different scenarios.

III. DESIGN REQUIREMENTS AND OVERALL DESIGN

In this section, the design requirements and overall design of the subway visualization system are described as following subsections, separately.

[1] Trains of data, http://senseable.mit.edu/trainsofdata/

A. SHANGHAI SUBWAY DATA

1) DATA DESCRIPTION

The datasets used in this system include the following three categories: Shanghai subway line and station data, Shanghai subway running data, and the smart card transaction data of Shanghai subway. In addition, the latitude and longitude of Shanghai subway stations we used in this system are from Baidu map. As of April 2015, Shanghai subway line has 617.53 kilometers and over 9 million passengers per day. The above datasets we used are multi-sourced information. Thus we can take more exploration and analysis on them. The detailed information of above three types of data is introduced as follows:

- **Shanghai subway line and station data.** This dataset contains 14 lines (including 289 stations) basic information of Shanghai subway in April 2015. All these data are stored in *.CSV* files, including multiple fields, such as the line number, line length of Shanghai subway.
- **Shanghai subway running data.** This dataset contains the train operations of Shanghai Subway Line 1 from July 2014 to April 2015. It records the number of train, the arrival and departure stations of the corresponding trains, the arrival and departure time of the corresponding trains, the time deviation of each train and other information. A total volume of this dataset is 800M.
- **The smart card transaction data of Shanghai subway.** This dataset includes Shanghai subway ferry data and part of the smart card transaction data of bus system from April 1, 2015, to April 30, 2015. The raw data are between 800M to 900M for each recorded day and with *.CSV* files. A total volume of this dataset is with 22 GB, more than 400 million pieces of data, including passenger's smart card number, passenger's transaction stations and the corresponding time, sum of consumption and other information.

2) DATA PREPROCESSING

The procedure of subway data preprocessing includes data cleaning, data merging and extraction of OD data. The data cleaning process includes processing missing data, inconsistent data and duplicated data. Before we visualize the subway data, we first clean the raw data to make sure the data we used are as percise as possible. Then, we merge the thirty days smart card transaction data (April 1, 2015 to April 30, 2015) into a collection for conducting the follow-up spatial-temporal explorations. In our system, we extract the OD data [27] from the smart card transaction data for station clustering module.

B. DESIGN REQUIREMENTS

For visualizing the subway data more efficiently and plainly, we decide to design a interactive system so that users can easily obtain the main subway information from the visualization charts and further analyze the laws of human mobility. Therefore, the demand of our system mainly contains four parts: 1) Provide the line and station data visualization schemes. 2) Enable users to conduct spatial-temporal exploration based on the smart card transaction data. 3) Provide station function exploration and analysis such that users can explore the function of different stations and the links among the different areas around these stations to further analyze the importance of different stations. We list the details of the above requirements as follows:

1) PROVIDE VISUALIZATION VIEWS BASED ON SUBWAY LINE AND STATION DATA

First of all, we consider the main problems such as how to show the time-varying law of subway data? How to identify the relation of subway lines? In order to solve such problems, we need to visualize the Shanghai subway basic data in an appropriate way. Therefore, we decided to visualize these subway data using three views: station line view, subway running view, and passenger flow view. The station line view shows the spatial distribution of the subway stations and line numbers. Besides, this view also can show the transfer stations. The subway running view shows each train's running visualization of Shanghai subway within one day such that users can have a macro perspective observation to the Subway Line 1. The passenger flow view shows the volume of passenger among different subway lines such that users can adjust the time range to observe the dynamic changes of passenger flow.

2) PROVIDE SPATIAL-TEMPORAL EXPLORATION BASED ON THE SMART CARD TRANSACTION DATA

Spatial and temporal attributes are the two basic characteristics of the smart card transaction data. In our system, the transaction data we used are abstract, with large-scale volume. Therefore, it is difficult for us to find the useful information from raw data without processing it. Thus, we propose an interactive interface which enables users to have both temporal and spatial exploration. On the basis of the above visual views of the subway basic data, it is necessary to provide the selection function (both spatial and temporal selection) in our visual system such that users can select the location, start time and end time to query the smart card transaction data interactively.

3) PROVIDE STATION FUNCTION EXPLORATION AND ANALYSIS

Based on the above subway data, we can build the time-dependent passenger flow model of subway stations by implementing clustering algorithm. Shanghai subway lines have 289 stations, which can be divided into several categories with different functions according to the similarity between passenger flow of these stations. Depending on the obtained categories, subway stations can be marked with different colors on the map. Finally, we display different passenger flows on charts.

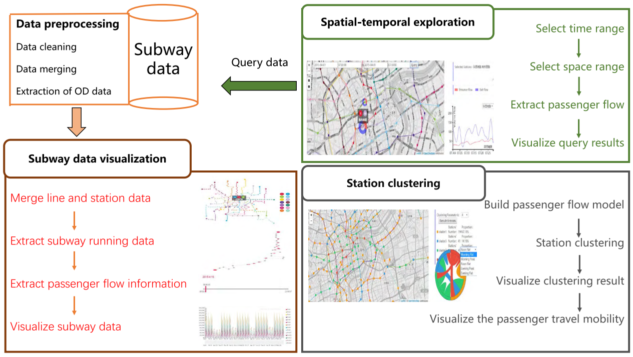

Data preprocessing
Data cleaning
Data merging
Extraction of OD data
Subway data
Query data
Spatial-temporal exploration
Select time range
Select space range
Extract passenger flow
Visualize query results
Subway data visualization
Merge line and station data
Extract subway running data
Extract passenger flow information
Visualize subway data
Station clustering
Build passenger flow model
Station clustering
Visualize clustering result
Visualize the passenger travel mobility

FIGURE 2. **Overall process design of the visualization system.**

C. OVERALL DESIGN

The overall process design of our visualization system is shown in Fig. 2. It consists of three modules: subway line and station data visualization module, spatial-temporal exploration visualization module, and station clustering visualization module. The detailed description of the above three modules is introduced in the following subsections, including the functional design and consideration of each module, the corresponding interface design, implementation details of the algorithm in these modules, and so on.

1) SUBWAY DATA VISUALIZATION MODULE

a: THE DESIGN AND IMPLEMENTATION OF MODULE FUNCTION

In this module, we visualize the preprocessed data from three views. First, we merge and match the line and station data (from the raw Shanghai subway line and station data) to prepare for the station line view. Then, we extract the time and location information of each running train from the subway running data to prepare for the subway running view. At last, we extract a whole month of passenger travel situation from the smart card transaction data of Shanghai subway to prepare for the passenger flow view.

In station line view, the dataset we used is the subway line data and detailed station data. We visualize these data using a map for showing the detailed information of subway stations and subway lines. When users click one station on the map. They can get the main information directly, including station name, the number of subway line, and whether the chosen station is a transfer station.

In the subway running view, the dataset we used is the subway running data. We visualize the subway running conditions of Line 1 within a day into a dynamic chart with a controllable timeline. After this process, users are able to interact with the timeline for observing the variation of subway running conditions with a day.

Passenger flow view is based on the smart card transaction data. In our system, each subway line is regarded as a unit, recording the volume of passenger flow among stations. Thus, we visualize the one-month smart card transaction data of Shanghai subway into 14 chordal graphs which correspond to 14 subway lines. After this process, users can interact with the time axis to inquire the specific volume of passenger flow for each subway line. This view also shows the variation of passenger flow within one month.

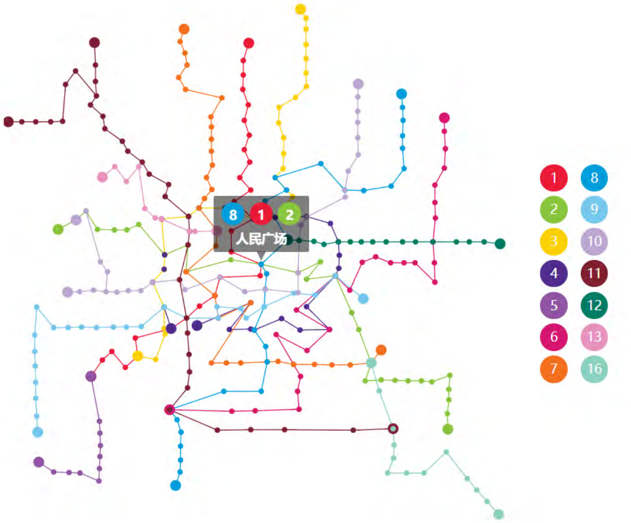


FIGURE 3. **The initial visualization interface of station line view.**

b: INTERFACE DESIGN OF SUBWAY DATA VISUALIZATION MODULE

In the subway data visualization module, station line view shows a total of 14 lines (Line 1 to Line 13, Line 16) of Shanghai subway, and the corresponding stations' information of each subway line, a total of 289 stations. The interface of station line view in our system is shown in Fig. 3, by selecting People's Square station. This interface can show the transfer information and the station name in the chart when users interact with the line chart. Additionally, the pie charts are associated with the number of subway line in the line chart. When users choose one or more than one subway lines, only the selected line chart will be displayed. This station line view can also help users understand the distribution of Shanghai subway lines and stations through an interactive way.

In the subway running view, we visualize the subway running data into an interactive line chart with the time axis. The data we used is selected from the raw running data about Shanghai subway Line 1 from 00 : 00 : 00 to 23 : 59 : 57 on April 16, 2015. The interface of the subway running view is shown in Fig. 4 and Fig. 5. Fig. 4 shows the "real-time" subway running view, where each running translucent red circle represents a train's position. Fig. 5 is the corresponding time axis for displaying the current time while trains are "running" in the line chart. Through this view, we can observe the train's operation within a day in an intuitive way.

As shown in Fig. 6, passenger flow view uses the multi-view linkage technology, in which the subject of the line chart is associated with the histogram on the right side. In the left line chart, the horizontal axis represents the time of passenger flow changes, the vertical axis shows the passenger

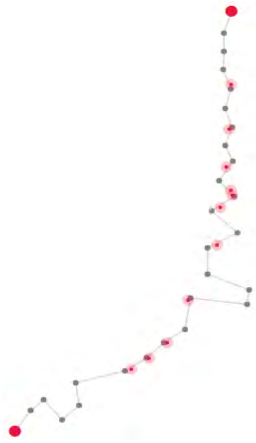


FIGURE 4. The visualization interface of the subway running view.

FIGURE 5. The corresponding time axis of the subway running view interface.

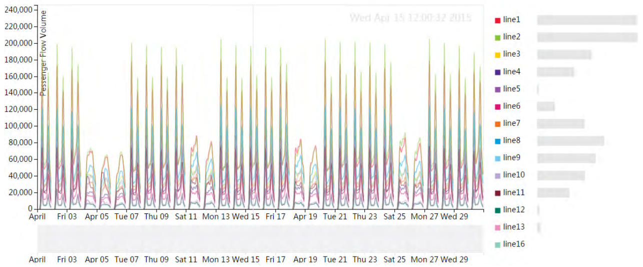


FIGURE 6. The initial visualization interface of passenger flow view. The scalable horizontal axis shows the time of passenger flow changes, users can select one specific time that the right chart will show the corresponding passenger flow volume of subway lines (14 lines correspond to 14 colors) on the vertical axis. The right histogram is also linked with the time axis, shows the specific volume of each subway lines, respectively.

flow volume, and the right histogram shows the specific passenger flow volume of each subway lines, respectively. When users interact with the line chart by selecting a specific time, the coordinated presentation of the two chart can not only show the specific information about passenger flows more intuitively, but also can enhance the understanding of the law of passenger flow changes in the subway data.

2) SPATIAL-TEMPORAL EXPLORATION MODULE

In this module, we limit the time range to one month and set the scope of space for selection function. According to the selected time and space, our system can query the preprocessed database and visualize the corresponding passenger travel case in the interface automatically.

a: THE DESIGN AND IMPLEMENTATION OF MODULE FUNCTION

The spatial-temporal exploration visualization module aims to process the large-scale and abstract smart card transaction data of Shanghai subway. Therefore, we provide the selection function based on the OpenStreet Map. Users can query the changes of traffic conditions by selecting a specific time range and geographical range. This module can visualize the query result using a ring graph and a line chart automatically. The ring graph shows the comparison of passenger flow volume between the selected stations while the line chart shows traffic changes over time.

b: INTERFACE DESIGN OF SPATIAL-TEMPORAL EXPLORATION MODULE

The interface of spatial-temporal exploration module before the query is shown in Fig. 7. This interface has four input boxes such that users can select the time range. Then, it enables users to select the geographical range on the map. After selecting time and geographical information and click on the ‘‘query’’ button, the query operation can be executed. Based on this, we divide the Spatial-temporal query operation into three steps: First, we use the station selection tool, circle the station users' want to explore. Then we specify the time range. Finally click the ‘‘query’’ button.

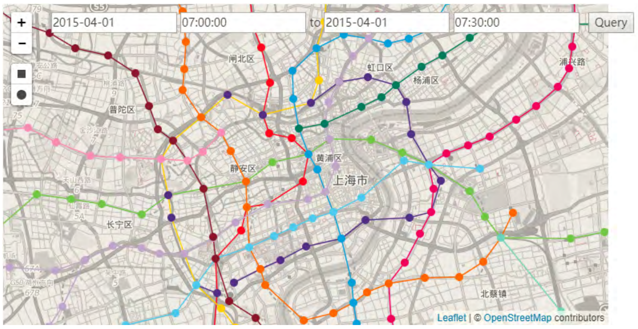
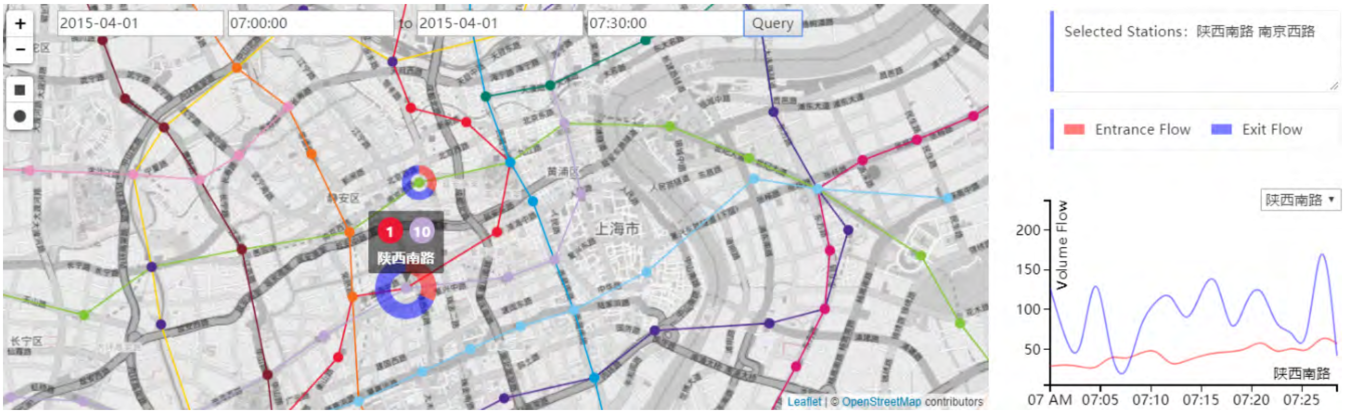


FIGURE 7. The initial interface of spatial-temporal exploration module. Where the four input boxes represent: start date, start time, end date, end time, respectively.

The result of the spatial-temporal query will be presented in two ways: First, this visualization module uses a doughnut chart to show the selected stations on the map, of which the radius of the doughnut represents the size of the corresponding passenger flow volume, while the red part represents the inbound passenger flow and the blue part represents the outbound passenger flow. Second, this module also uses a line chart for showing the time-dependent passenger flow changes of each selected stations, respectively. For instance, we select two stations: Shanxi South Road and Nanjing West Road, then, we set the time range from 7 : 00 to 7 : 30 on April 1, 2015. The query result is shown in Fig. 8, we can find this design shows the location of the selected stations and their passenger flow comparison in an intuitive way that



2015-04-01 07:00:00 to 2015-04-01 07:30:00 Query

Selected Stations: 陕西南路 南京西路

FIGURE 8. The query result of spatial-temporal exploration module. The right doughnut chart shows the entrance (red) and exit (blue) flow volume of the selected Nanjing West Road. The left line chart shows the time-dependent entrance and exit flow changes of the selected Shanxi South Road.

users can complete the spatial-temporal exploration through this module.

3) STATION CLUSTERING MODULE

In this module, we build the passenger flow model and cluster the stations based on the basis of passenger flow volume, then we visualize the clustering results. After station clustering, the stations with similar function are divided into several categories, respectively.

a: THE DESIGN AND IMPLEMENTATION OF MODULE FUNCTION

In station clustering module, we divide each day into several travel peak segments according to the law of passenger flow in Shanghai subway lines. Then, based on these peaks, we extract the passenger flow model of each station. Finally, we identify the main function of each station through the extracted passenger flow models by implementing $k-means$ clustering algorithm, the 289 stations are divided into several categories with the similar function (passenger flow). At this time, we embed this process into our station clustering module and combine with spatial-temporal exploration module, aims to provide users an intuitive interface to display the clustering results that they can find the stations which are with the similar function, analyze the characteristics of passenger flow between various stations, furthermore, analyze the functional characteristics and outliers.

We introduce the steps before clustering stations about dividing passenger flow peak and extracting passenger flow model of each station in the following.

- **Divide passenger flow peak.** According to the law of time-dependent passenger flow, we divide a day into five periods: morning flat, morning peak, noon flat, evening peak, evening flat, the specific period partition of a day is shown in the TABLE 1. Passengers' travel by subway in the same condition are generally similar, for instance, when passenger flow appears evening peak on a workday, the passengers usually travel from the stations near the workplaces to stations near the residence or traffic center.

TABLE 1. Period partition of a day.

Time period	Start time	End time
Morning peak	05:00:00	07:00:00
Morning flat	07:00:00	09:00:00
Noon flat	09:00:00	17:00:00
Evening peak	17:00:00	19:30:00
Evening flat	19:30:00	00:00:00

- **Extract passenger flow model of each station.** First, we treat each trip of passengers as a unit, thus we have a total of M trip units. Based on this, we can extract the travel model $F = (o, d, p)$, where o represents the originating station of the trip, d represents the destination station of the trip, p denotes the period of the corresponding passenger flow. Thereby, we obtain M travel models F. Then, we extract the passenger flow model of each station. Here, we assume each subway station can represent the main function of its surrounding area and show the different passenger flow rule in different travel periods. Therefore, we define the total number of stations as T, the number of the time period of the corresponding station as N. For each station, each time period corresponds to inbound and outbound passenger flow. Based on this, we can build the passenger flow model matrix as $S = T^*(2^*N)$, where each row of matrix S represents one station's passenger flow pattern. The matrix has a total of 2^*N columns, representing the N time periods' inbound and outbound passenger flow.

b: STATION PARTITION BASED ON K-MEANS CLUSTERING ALGORITHM

Station partition is based on the smart card transaction data of Shanghai subway. After modeling the characteristics of subway station's passenger flow and extracting passenger flow of each station, the interactive interface of this clustering module enables users to select the clustering algorithm parameters for clustering the subway stations. The procedures are introduced as follows:

- The division of time dimension. According to the operation time of Shanghai subway and the pattern of passenger travel regularity, the passenger flow of each station is divided into five periods within a day. The specific partition is shown in Table 1.
- The division of spatial dimension. According to the Shanghai subway basic data described above, we take each station as a unit. Therefore, all stations are divided into 289 spatial units.
- For each spatial unit, we count the passenger's inbound and outbound number in the five period segments respectively, then, a 289*8 passenger characteristic matrix D is formed. This matrix D is obtained from the stations' passenger flow model.
- Receive the input parameter $k(k \leq n)$ dynamically and take this parameter as the cluster number of k-means clustering algorithm.
- We randomly select k objects from the matrix D as the centers of the k clusters. The centers of the k clusters after initialization are denoted in Equ. 1.

$$\mu^{(0)} = \mu_1^{(0)}, \ldots, \mu_k^{(0)} \tag{1}$$

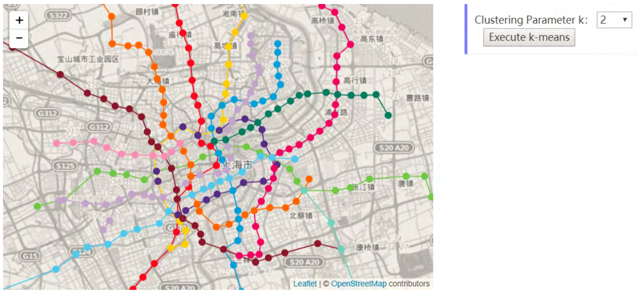

Clustering Parameter k: 2
Execute k-means

FIGURE 9. The initial interface of the station clustering module.

- Then, as shown in Equ. 2, we calculate the similarity between each object $j(j \in 1, \ldots, n)$ in k clusters and the center of the k clusters, and divide these objects into clusters of the closest center point respectively.

$$Cluster^{(t)}(j) \longleftarrow \arg\min_i \|\mu_i - x_j\|^2 \tag{2}$$

- Recalculate the center of k clusters according to the clustering results. The formula is shown in Equ. 3

$$\mu_i^{(t+1)} \longleftarrow \arg\min_{\mu} \sum_{j:Cluster(j)=i} \|\mu - x_j\|^2 \tag{3}$$

- Repeat the two steps in Equ. 2 and Equ. 3, until the center of each cluster remains no longer changes. At this time, we complete the station clustering.

c: INTERFACE DESIGN OF STATION CLUSTERING MODULE

The initial interface of the clustering module is shown in Fig. 9. It consists of three parts: the map view, the cluster parameter selection box and the execute button for implementing k-means clustering algorithm. The map view displays the original relationship between 14 lines of Shanghai subways. The process of station clustering is as follows: First, select the parameter k, representing the number of classes to be clustered. Then click the "execute k-means" button to obtain the clustering result automatically. As shown in Fig. 10, the interface after clustering shows the clustering results on the left map in different colors, where the same color represent the stations are with the same function. The right side of this interface shows the detailed categories of clustering results, the legends of the clusters and their corresponding proportion. The volume of passenger flow between different stations within the same time period (morning flat) is shown in a chord chart in the lower right of this interface.

In the chord chart, according to the above time partition, passenger flow of different stations are presented within five time periods respectively. The width of the chord represents the volume of passenger flow. When one chord is selected,

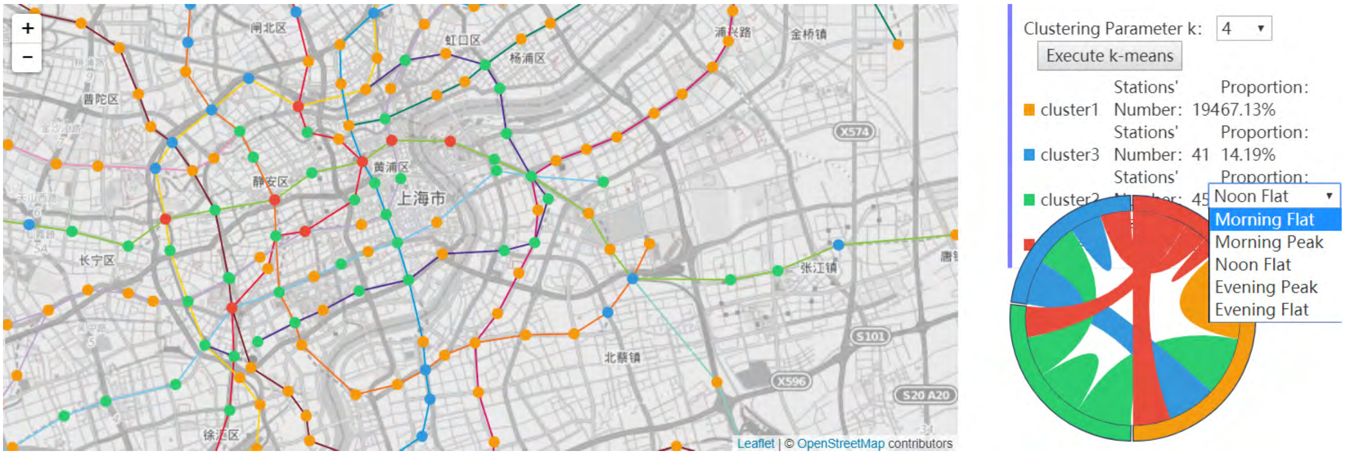

Clustering Parameter k: 4
Execute k-means
🟧 cluster1 Stations' Number: 194 Proportion: 67.13%
🟦 cluster3 Stations' Number: 41 Proportion: 14.19%
🟩 cluster2 Stations' Number: 45 Proportion:

FIGURE 10. Interface of the station clustering module after clustering.

the specific passenger flow of the two clusters will be displayed.

IV. VISUAL ANALYSIS BASED ON VISOS

In this section, we take the visualization analysis in three aspects, including visualization analysis of human mobility, visualization analysis of spatial-temporal exploration and visualization analysis of station clustering results.

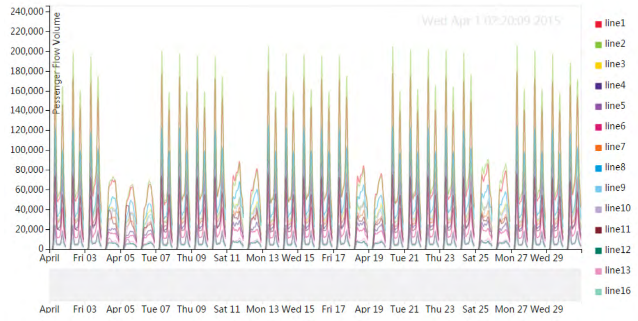


FIGURE 11. **The initial interface of passenger flow view, of which the time range is limited to one month.**

A. VISUAL ANALYSIS OF HUMAN MOBILITY

We analyze the human mobility based on the volume changes of passenger flow, of which the passenger flows are extracted from the smart card transaction data. This dataset contains a whole month of passenger travel records. Thus, we can take the analysis of human mobility on the passenger flow changes. Fig. 11 shows the volume changes of passenger flow in April 2015. From this figure, we can find that the passenger flows of Shanghai subway repeatedly changes in April. Furthermore, we can find the passenger flows form four complete cycles by combining the time axis. These four cycles correspond to four weeks within a month. After zooming in the time axis, as shown in Fig. 12, we limit the time range to one week, from April 13 to April 19. It is clearly for us to find that the weekly passenger flows consist of five high peaks and two low peaks. Observe the time axis in each cycle of Fig. 11, we can find the five high peaks exactly correspond to workdays (Monday to Friday), and the low peaks correspond to the weekend. The exception is the first cycle, which will be explained later. The similar conclusion can be drawn from other cycles. On workdays, the commuter group mainly forms the passenger flow, passenger flows are far greater than the weekends. Thus, the reason leads to the high peak and low peak.

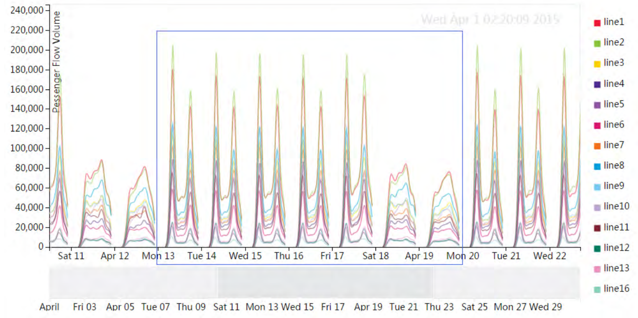


FIGURE 12. **The line chart within this blue wireframe shows the passenger flow from April. 13 to April. 19.**

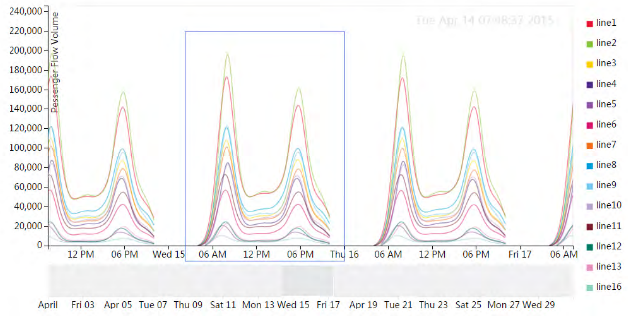


FIGURE 13. **The line chart within this blue wireframe shows the passenger flow in April. 15.**

Then, we continue zooming in the view as shown in Fig. 13, we can find the breakpoint on the time axis. From the occurrence of these breakpoints on the horizontal axis, we can find that all of these breakpoints have appeared at night due to night outage. Furthermore, we can also find there are two passenger flow peaks in one workday. By observing the time under these two peaks, we can find these peaks are around $08:00 a.m.$ and $06:00 p.m.$, which means the morning peak and evening peak as we mentioned above.

The content described above is the passenger flow periodically changed in seven days, one cycle consists of five workdays and two-day weekends. While in the first cycle, the weekly passenger flows consist of four high peak days and three low peak days. By observing the corresponding low peak days' timeline, we find the date of the three low peak days corresponds to April 4, 5, 6 in 2015. We assume that the public holidays may be similar traffic condition to the weekends. After querying the lunar calendar, we find these three days are the Chinese Ching Ming Festival.

B. VISUALIZATION ANALYSIS OF SPATIAL-TEMPORAL EXPLORATION

Based on the spatial-temporal exploration module, we can query the inbound and outbound passenger flow changes of different stations within five weekdays. Compared to the different passenger flows, we can find two typical characteristics of subway stations, that is: morning peak-led station and evening peak-led station. We take the comparison of Jinjiang Park Station and Century Avenue Station as examples for illustrating the above two characteristics.

As shown in Fig. 14, the two charts are passenger flow change curves of Jinjiang Park Station and Century Avenue Station during the workdays with the morning peak and evening peak. Fig. 14(a) shows the inbound passenger flow is significantly more higher than the outbound passenger flow

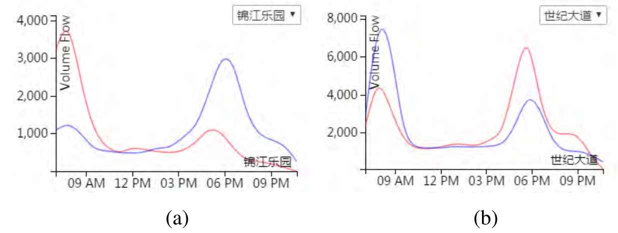


FIGURE 14. Passenger flow line chart. Of which the blue curve represents the outbound passenger flow, the red curve represents the inbound passenger flow. (a) Passenger flow of Jinjiang Park Station within a day. (b) Passenger flow of Century Avenue Station within the same day.

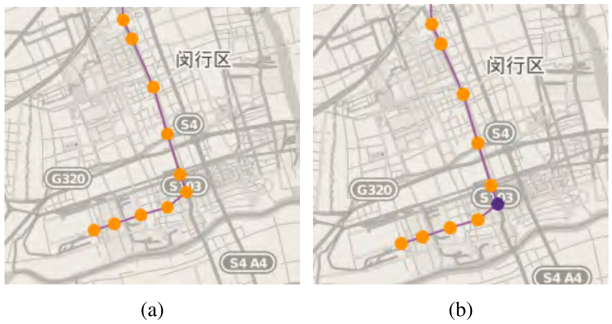


FIGURE 15. Comparison of the clustering result while the clustering parameter changed to 4 and 5, respectively. (a) Part stations of the clustering result in Line 5 when the parameter is set to 4. (b) Part stations of the clustering result in Line 5 when the parameter is set to 5.

during the morning peak, but the outbound passenger flow is much more higher than the inbound passenger flow during the evening peak. On the contrary, Fig. 14(b) shows a totally opposite passenger flow changes comparing to Fig. 14(a). There are a series of stations with the similar passenger flow to Jinjiang Park and Century Avenue. We name the stations similar to Jinjiang Park as morning inbound evening outbound passenger flow mode. Similarly, we name the stations to Century Avenue as morning outbound evening inbound passenger flow mode. Because most of the passenger flow generated in the workdays are commuters, the stations of the morning inbound evening outbound passenger flow mode are nearby the residential area, and the stations of the morning outbound evening inbound passenger flow mode are nearby the workplace.

C. VISUALIZATION ANALYSIS OF CLUSTERING RESULTS

As shown in Fig. 15, we make a comparison between the same part of stations by changing the clustering parameter from 4 to 5. When the parameter is set to 4, the stations shown in Fig. 15(a) are clustered to the same station with the similar function. While the clustering parameter is set to 5, an outlier station appears (the purple station). From this phenomenon, we deduce that this special station's traffic model is different from the rest of other stations. In order to verify this speculation, we return back to the spatial-temporal exploration visualization module, select the same spatial range of Fig. 15, and limit the time range to one day, the query result is shown in Fig. 16. We can clearly observe that passenger flow of the spacial station is much larger than the surrounding other stations, which proves the applicability of our visualization system.

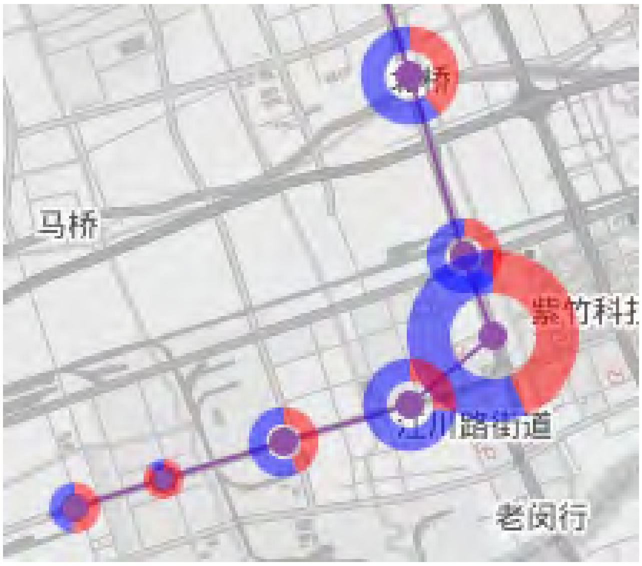


FIGURE 16. Spatial-temporal query on part of Line 5.

V. CONCLUSION

In this paper, we present a visual analysis system named VISOS based on Shanghai subway datasets. VISOS consists of three modules: the subway data visualization module, the spatial-temporal exploration module, and the station clustering module. We aim at displaying the station's distribution and passenger flow changes with spatial and temporal attributes by mining the raw subway data. Furthermore, we enable the users to explore the human mobility by conducting a series of visual analysis based on our system. In VISOS, the datasets we obtained are from the Shanghai subway. Despite we can not access the other cities' subway data yet, the subway data are with the similarity of time and location attributes, and time. In other words, we can process the other cities' subway data in the same workflow and take spatial-temporal exploration and analysis based on them.

REFERENCES

[1] F. Xia, J. Wang, X. Kong, Z. Wang, J. Li, and C. Liu, "Exploring human mobility patterns in urban scenarios: A trajectory data perspective," *IEEE Commun. Mag.*, vol. 56, no. 3, pp. 142–149, Mar. 2018.

[2] T. Tang *et al.*, "FISS: Function identification of subway stations based on semantics mining and functional clustering," *IET Intell. Transport Syst.*, 2018, doi: 10.1049/iet-its.2017.0316.

[3] Z. Ning, F. Xia, N. Ullah, X. J. Kong, and X. P. Hu, "Vehicular social networks: Enabling smart mobility," *IEEE Commun. Mag.*, vol. 55, no. 5, pp. 16–55, May 2017.

[4] J. Wang, X. Kong, A. Rahim, F. Xia, A. Tolba, and Z. Al-Makhadmeh, "IS2Fun: Identification of subway station functions using massive urban data," *IEEE Access*, vol. 5, pp. 27103–27113, 2017.

[5] F. Ghofrani, Q. He, R. M. P. Goverde, and X. Liu, "Recent applications of big data analytics in railway transportation systems: A survey," *Transp. Res. C, Emerg. Technol.*, vol. 90, pp. 226–246, May 2018.

[6] W. Chen, F. Guo, and F. Y. Wang, "A survey of traffic data visualization," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 6, pp. 2970–2984, Jun. 2015.

[7] T. Gonçalves, A. P. Afonso, and B. Martins, "Cartographic visualization of human trajectory data: Overview and analysis," *J. Location Based Services*, vol. 9, no. 2, pp. 138–166, 2015.

[8] Y. Zheng, W. Wu, Y. Chen, H. Qu, and L. M. Ni, "Visual analytics in urban computing: An overview," *IEEE Trans. Big Data*, vol. 2, no. 3, pp. 276–296, Sep. 2016.

[9] W. Zeng, C. W. Fu, S. M. Arisona, A. Erath, and H. Qu, "Visualizing mobility of public transportation system," *IEEE Trans. Vis. Comput. Graphics*, vol. 20, no. 12, pp. 1833–1842, Dec. 2014.

[10] Y. Zheng, "Trajectory data mining: An overview," *ACM Trans. Intell. Syst. Technol.*, vol. 6, no. 3, p. 29, 2015.

[11] A. Rahim *et al.*, "Vehicular social networks: A survey," *Pervasive Mobile Comput.*, vol. 43, pp. 96–113, Jan. 2018.

[12] X. Li, G. Li, F. Yang, J. Teng, D. Xuan, and B. Chen, "Traffic at-a-glance: Time-bounded analytics on large visual traffic data," in *Proc. IEEE Int. Conf. Comput. Commun. (IEEE INFOCOM)*, Apr. 2016, pp. 1–9.

[13] S. Chen *et al.*, "Interactive visual discovering of movement patterns from sparsely sampled geo-tagged social media data," *IEEE Trans. Vis. Comput. Graphics*, vol. 22, no. 1, pp. 270–279, Jan. 2016.

[14] X. Kong, X. Song, F. Xia, H. Guo, J. Wang, and A. Tolba, "LoTAD: Long-term traffic anomaly detection based on crowdsourced bus trajectory data," *World Wide Web*, vol. 21, no. 3, pp. 825–847, 2017.

[15] W. Wu *et al.*, "TelCoVis: Visual exploration of co-occurrence in urban human mobility based on telco data," *IEEE Trans. Vis. Comput. Graphics*, vol. 22, no. 1, pp. 935–944, Jan. 2016.

[16] F. Xia, A. Rahim, X. Kong, M. Wang, Y. Cai, and J. Wang, "Modeling and analysis of large-scale urban mobility for green transportation," *IEEE Trans. Ind. Informat.*, vol. 14, no. 4, pp. 1469–1481, Apr. 2018.

[17] Z. Ning, X. Wang, X. Kong, and W. Hou, "A social-aware group formation framework for information diffusion in narrowband Internet of Things," *IEEE Internet Things J.*, vol. 5, no. 3, pp. 1527–1538, Jun. 2018.

[18] S. Al-Dohuki *et al.*, "SemanticTraj: A new approach to interacting with massive taxi trajectories," *IEEE Trans. Vis. Comput. Graphics*, vol. 23, no. 1, pp. 11–20, Jan. 2017.

[19] X. Kong *et al.*, "Mobility dataset generation for vehicular social networks based on floating car data," *IEEE Trans. Veh. Technol.*, vol. 67, no. 5, pp. 3874–3886, May 2018.

[20] N. Andrienko and G. Andrienko, "Visual analytics of movement: An overview of methods, tools and procedures," *Inf. Vis.*, vol. 12, no. 1, pp. 3–24, 2012.

[21] W. Cui, H. Zhou, H. Qu, P. C. Wong, and X. Li, "Geometry-based edge clustering for graph visualization," *IEEE Trans. Vis. Comput. Graphics*, vol. 14, no. 6, pp. 1277–1284, Oct. 2008.

[22] C. Tominski, H. Schumann, G. Andrienko, and N. Andrienko, "Stacking-based visualization of trajectory attribute data," *IEEE Trans. Vis. Comput. Graphics*, vol. 18, no. 12, pp. 2565–2574, Dec. 2012.

[23] N. Ferreira, J. Poco, H. T. Vo, J. Freire, and C. T. Silva, "Visual exploration of big spatio-temporal urban data: A study of New York City taxi trips," *IEEE Trans. Vis. Comput. Graphics*, vol. 19, no. 12, pp. 2149–2158, Dec. 2013.

[24] X. Huang, Y. Zhao, J. Yang, C. Zhang, C. Ma, and X. Ye, "Trajgraph: A graph-based visual analytics approach to studying urban network centralities using taxi trajectory data," *IEEE Trans. Vis. Comput. Graphics*, vol. 22, no. 1, pp. 160–169, Jan. 2016.

[25] Y. Yang, T. Dwyer, S. Goodwin, and K. Marriott, "Many-to-many geographically-embedded flow visualisation: An evaluation," *IEEE Trans. Vis. Comput. Graphics*, vol. 23, no. 1, pp. 411–420, Jan. 2017.

[26] F. Miranda *et al.*, "Urban pulse: Capturing the rhythm of cities," *IEEE Trans. Vis. Comput. Graphics*, vol. 23, no. 1, pp. 791–800, Jan. 2017.

[27] D. Guo and X. Zhu, "Origin-destination flow data smoothing and mapping," *IEEE Trans. Vis. Comput. Graphics*, vol. 20, no. 12, pp. 2043–2052, Dec. 2014.

TAO TANG is currently pursuing the bachelor's degree in communication and information engineering with the Chengdu College, University of Electronic Science and Technology of China, Chengdu, China. His research interests include big data analytics and visualization.

XIANGJIE KONG (M'13–SM'17) received the B.Sc. and Ph.D. degrees from Zhejiang University, Hangzhou, China. He is currently an Associate Professor with the School of Software, Dalian University of Technology, China. He has authored over 70 scientific papers in international journals and conferences (with 50+ indexed by ISI SCIE). His research interests include intelligent transportation systems, mobile computing, and cyber-physical systems. He is a Senior Member of CCF and a member of ACM. He served as the (Guest) Editor for several international journals and the Workshop Chair or a PC Member for a number of conferences.

MENGLIN LI received the bachelor's degree in software engineering from the Dalian University of Technology, Dalian, China, in 2016. She is currently pursuing the master's degree with the Alpha Lab, School of Software, Dalian University of Technology. Her research interests include big traffic data mining and analysis, human mobility behavior analysis, and smart city development.

JINZHONG WANG received the B.Sc. degree in computer education from Anshan Normal University, Anshan, China, in 2002, and the M.Sc. degree in computer application technology from Liaoning University, Shenyang, China, in 2005. He is currently pursuing the Ph.D. degree with the School of Software, Dalian University of Technology, Dalian, China. Since 2005, he has been with Shenyang Sport University, Shenyang. His research interests include computational social network, network science, data science, and mobile social networks.

GUOJIANG SHEN received the B.Sc. degree in control theory and control engineering and the Ph.D. degree in control science and engineering from Zhejiang University, Hangzhou, China, in 1999 and 2004, respectively. He is currently a Professor with the College of Computer Science and Technology, Zhejiang University of Technology. His current research interests include intelligent control theory and application, advanced control technology and application, and urban road traffic modeling and control technology.

XINSHUANG WANG received the M.Sc. degree from the School of Software Technology, Dalian University of Technology, Dalian, China, in 2017. Since 2017, she has been a Front-End Developer with Tencent. Her research interests are data visualization.